



BUILD ['23]

BEST OF - KOREA

The Data Cloud Dev Summit

GC 데이터 플랫폼 구축 사례

Snowflake를 이용한 시각화 프로젝트

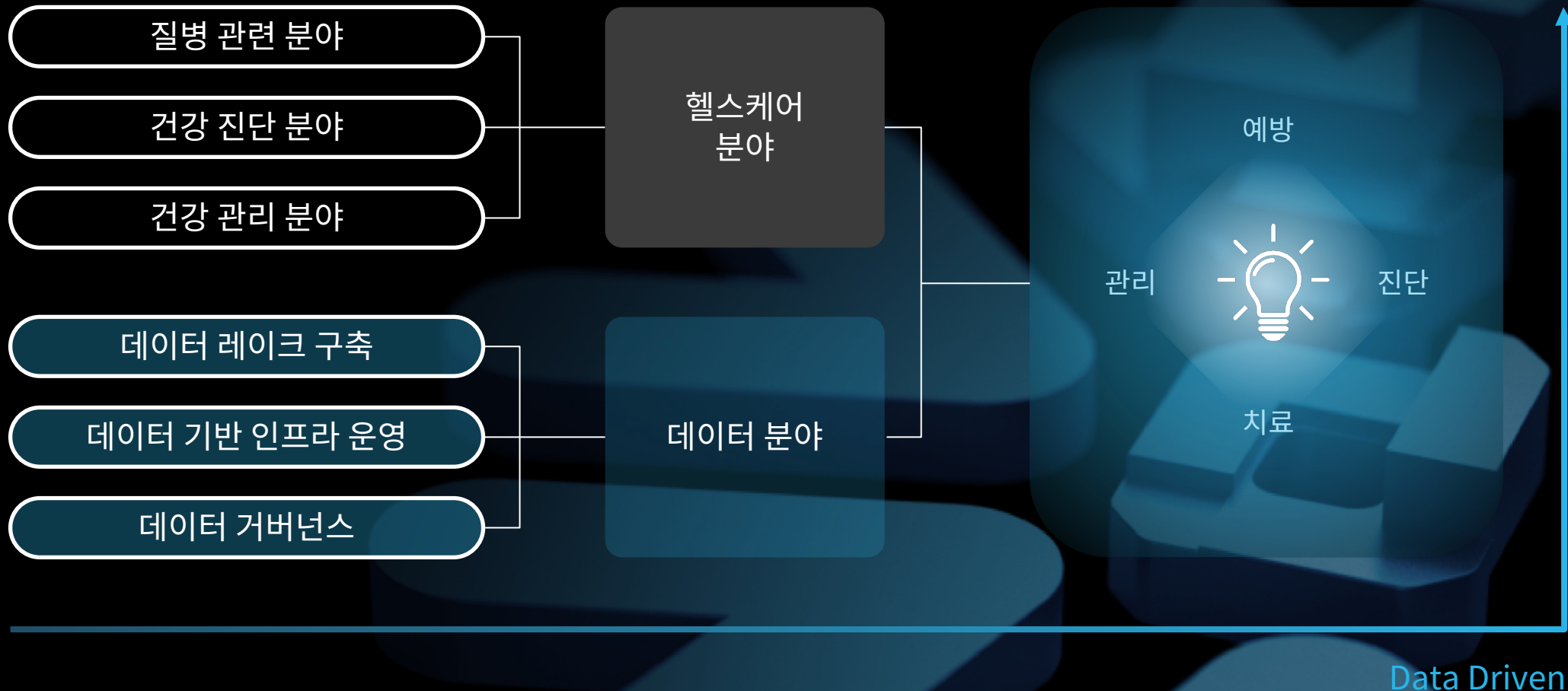
녹십자 홀딩스 Dx담당
조예진 프로

[GC 소개]

“Total Healthcare Solution 기업 추구”



[Data Driven Healthcare]



[Why Snowflake?]

방향성

- HW/SW의 구축 최소화
- 장기적인 관점에서 다양한 CSP 지원
- 데이터 레이크, BI서비스 독립구축

유연성

- 최소한의 리소스로 비용 절약
- Scale-Up/Out 용이
- 진입장벽이 낮아 적은 인원으로 구축가능

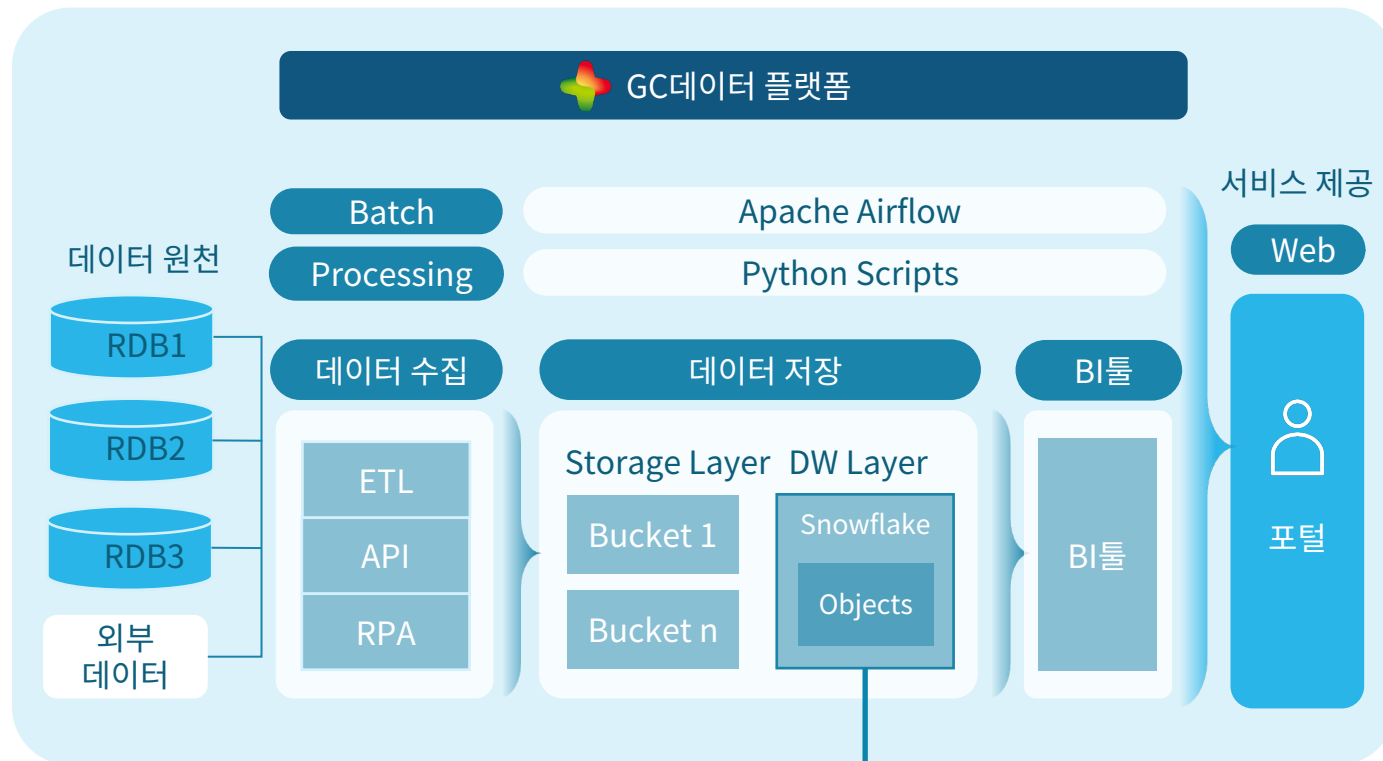
효율성

- 클러스터링 성능 부분 자동 지원
- 파일기반의 분산처리로 대용량 빠르게 처리

운영성

- 다양한 References 존재
- 버퍼역할을 하는 대용량 Cache영역 제공

[데이터플랫폼 아키텍처]



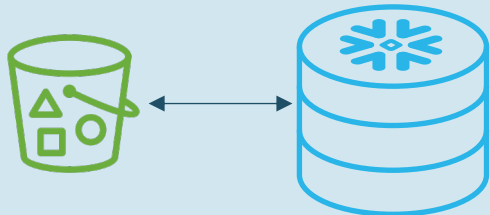
[Lesson Learn 1. 사용성 측면의 구축]

편의성 및 자유도가 높은 데이터 적재 및 로드 방식 제공



1. 쉬운 데이터로드

클라우드 스토리지의 파일이 테이블화
되어 적재 전 SQL연산 가능

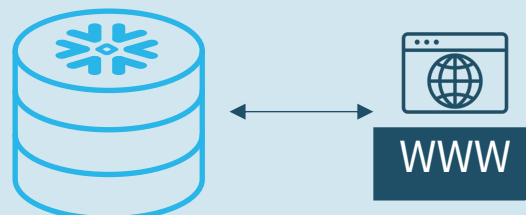


편의성 제공



2. 유연한 데이터 인터페이스

SQL REST API 방식으로
전사적 협업 가능



클라이언트 자유도 제공



3. SQL의 단순화

인덱스 튜닝 불필요
튜닝을 최소화하여 DBA이 부담 경감



튜닝 최소화

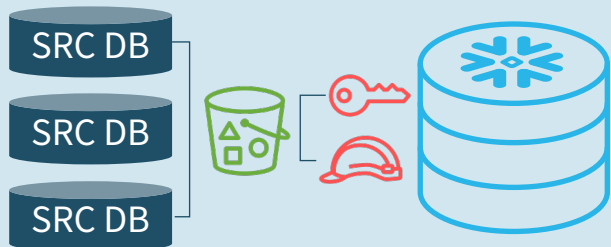
[Lesson Learn 2. 보안성 측면의 구축]

Snowflake만으로 수준 높은 보안이 적용된 데이터 플랫폼 구축



1. Role 방식 vs Key 방식

신뢰관계를 이용한 Stage 생성

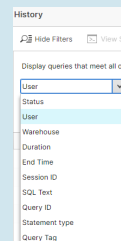


Role 방식 채택



2. 보안 감시

개별적 권한 부여 및 모든 쿼리 기록



모든 쿼리 모니터링



3. 쉬운 비식별처리

Dynamic Data Masking기능으로 다른 솔루션을 도입하지 않고도 비식별처리



SQL로 편하게 설정

[Lesson Learn 3. 활용성 측면의 구축]

BI툴과 연동하는데 이로운 3가지 포인트 존재



1. 빠른 적재

라이브 커넥터, 인메모리큐브 지원



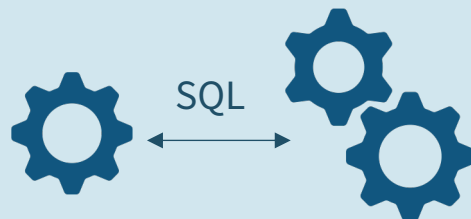
BI툴

BI툴과 연동 용이



2. 디커플링의 편의성

Warehouse 쉽게 성능 고도화

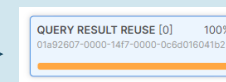
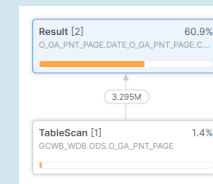


성능 쉽게 변경 가능



3. 버퍼역할의 대용량 Cache

클라우드 스토리지가 Cache 영역으로 BI툴과 연동 용이



파라미터 조정 불필요

[Future Works]

사용 확산 측면



- 전사적 사용 확산
- 서로 간의 **Data Sharing**

정합성 측면



- Hybrid Table 변경
- 데이터 정합성을 유지하기 위한 노력 최소화

민첩성 측면



- 파이프라인 단순화
- Snowpipe를 이용한 실시간 데이터 플랫폼

감사합니다!

yejinjo@gccorp.com